

What is claimed is:

1. A method for comparing one or more nucleic acids in a first sample comprising a first population of first nucleic acids having different nucleotide sequences, with a second sample comprising a second population of second nucleic acids having different nucleotide sequences, wherein the different nucleotide sequences of said second population are known, said method comprising:

(a) probing said first sample with one or more recognition means, each recognition means recognizing a different target nucleotide subsequence or a different set of target nucleotide subsequences;

(b) generating one or more output signals from said first sample probed by said recognition means, each output signal being produced from a first nucleic acid in said sample by recognition of one or more target nucleotide subsequences in said first nucleic acid by said recognition means and comprising a representation of (i) the length between occurrences of target nucleotide subsequences in said first nucleic acid, and (ii) the identities of said target nucleotide subsequences in said first nucleic acid or the identities of said sets of target nucleotide subsequences among which are included the target nucleotide subsequences in said first nucleic acid; and

(c) comparing each representation provided by said first sample with the different nucleotide sequences of said second population by generating a statistical score which states the probability that one or more output signals from said first sample indicates that one or more nucleotide sequences of said first sample are either present or absent in said second sample.

2. The method of claim 1, wherein the output signals from said first sample indicate that one or more nucleotide sequences of said first sample are present in said second sample.

09900624.070601

3. The method of claim 1, wherein the output signals from said first sample indicate that one or more nucleotide sequences of said first sample are absent in said second sample.
4. The method of claim 1, wherein said statistical score is generated according to Bayes Rule of conditional probability.
5. The method of claim 1, wherein said length between occurrences of target nucleotide subsequences is pre-determined.
6. The method of claim 1, wherein said comparison is performed by additionally generating a statistical score of the false-negative rate, said false-negative rate being a representation of the probability that said output from said first nucleic acid or said second nucleic acid is not generated given that said first or second nucleic acid is present.
7. The method of claim 1, wherein said comparison is performed by additionally generating a statistical score of the false-positive rate, said false-positive rate being a representation of the probability that said output from said first nucleic acid or said second nucleic acid is generated given that said first or second nucleic acid is not present.
8. A method for comparing one or more nucleic acids in a first sample comprising a first population of first nucleic acids having different nucleotide sequences, with a second sample comprising a second population of second nucleic acids having different nucleotide sequences, wherein the different nucleotide sequences of said first and second samples are not known, said method comprising:
 - (a) probing said first sample with one or more recognition means, each recognition means recognizing a different target nucleotide subsequence or a different set of target nucleotide subsequences;
 - (b) generating one or more output signals from said first sample probed by said recognition means, each output signal being produced from a nucleic acid in said sample by recognition of one or more target nucleotide subsequences in said nucleic acid by said recognition means and comprising a representation of (i) the length

between occurrences of target nucleotide subsequences in said nucleic acid, and (ii) the identities of said target nucleotide subsequences in said first nucleic acid or the identities of said sets of target nucleotide subsequences among which are included the target nucleotide subsequences in said first nucleic acid;

(c) performing steps (a) and (b) for said second sample; and

(d) comparing the representation provided by said first sample with the representation provided by the second sample by generating a statistical score which states the probability that one or more output signals from said first sample indicates that one or more nucleotide sequences of said first sample are either present or absent in said second sample.

9. The method of claim 8, wherein the output signals from said first sample indicate that one or more nucleotide sequences of said first sample are present in said second sample.
10. The method of claim 8, wherein the output signals from said first sample indicate that one or more nucleotide sequences of said first sample are absent in said second sample.
11. The method of claim 8, wherein said statistical score is generated according to Bayes Rule of conditional probability.
12. The method of claim 8, wherein said statistical score is generated by:
 - i) performing steps (a) through (d) on two or more independent samples of said first and said second sample;
 - ii) generating an average representation of said independent samples; and
 - iii) generating a p-value from an f-test performed on said average representation.
13. The method of claim 12, wherein said p-value is between about 0.05 and about 0.10.
14. The method of claim 12, wherein said length between occurrences of target nucleotide subsequences is pre-determined.

15. The method of claim 8, wherein said length between occurrences of target nucleotide subsequences is pre-determined.
16. The method of claim 8, wherein said comparison is performed by additionally generating a statistical score of the false-negative rate, said false-negative rate being a representation of the probability that said output from said first nucleic acid or said second nucleic acid is not generated given that said first or second nucleic acid is present.
17. The method of claim 8, wherein said comparison is performed by additionally generating a statistical score of the false-positive rate, said false-positive rate being a representation of the probability that said output from said first nucleic acid or said second nucleic acid is generated given that said first or second nucleic acid is not present.
- ~~18.~~ A method for identifying or classifying one or more nucleic acids in a first sample comprising a first population of first nucleic acids having different nucleotide sequences, said method comprising:
- (a) probing said first sample with one or more recognition means, each recognition means recognizing a different target nucleotide subsequence or a different set of target nucleotide subsequences;
- (b) generating one or more output signals from said first sample probed by said recognition means, each output signal being produced from a first nucleic acid in said sample by recognition of one or more target nucleotide subsequences in said first nucleic acid by said recognition means and comprising a representation of (i) the length between occurrences of target nucleotide subsequences in said first nucleic acid, and (ii) the identities of said target nucleotide subsequences in said first nucleic acid or the identities of said sets of target nucleotide subsequences among which are included the target nucleotide subsequences in said first nucleic acid; and
- (c) comparing each representation provided by said first sample with a second population of second nucleic acids having different nucleotide sequences, wherein the different nucleotide sequences of said second population are known, by generating a

statistical score which states the probability that one or more output signals from said first sample indicates that one or more nucleic acid sequences of said first sample are either present or absent in said second sample;

whereby said indication of presence or absence provides said identification or classification.

19. The method of claim 18, wherein the output signals from said first sample indicate that one or more nucleotide sequences of said first sample are present in said second sample.

20. The method of claim 18, wherein the output signals from said first sample indicate that one or more nucleotide sequences of said first sample are absent in said second sample.

21. A method for identifying or classifying one or more nucleic acids in a first sample comprising a first population of first nucleic acids having different nucleotide sequences, said method comprising:

(a) probing said first sample with one or more recognition means, each recognition means recognizing a different target nucleotide subsequence or a different set of target nucleotide subsequences;

(b) generating one or more output signals from said first sample probed by said recognition means, each output signal being produced from a first nucleic acid in said sample by recognition of one or more target nucleotide subsequences in said first nucleic acid by said recognition means and comprising a representation of (i) the length between occurrences of target nucleotide subsequences in said first nucleic acid, and (ii) the identities of said target nucleotide subsequences in said first nucleic acid or the identities of said sets of target nucleotide subsequences among which are included the target nucleotide subsequences in said first nucleic acid;

(c) performing steps (a) and (b) for a second population of second nucleic acids having different nucleotide sequences, wherein the different nucleotide sequences of said second population are not known;

(d) comparing each representation provided by said first sample with said second sample, by generating a statistical score which states the probability that one or more output signals from said first sample indicates that one or more nucleotide sequences of said first sample are either present or absent in said second sample;

whereby said indication of presence or absence provides said identification or classification.

22. The method of claim 21, wherein the output signals from said first sample indicate that one or more nucleotide sequences of said first sample are present in said second sample.

23. The method of claim 21, wherein the output signals from said first sample indicate that one or more nucleotide sequences of said first sample are absent in said second sample.

24. A method for identifying, classifying, or quantifying one or more nucleic acids in a first sample comprising a first population of first nucleic acids having different nucleotide sequences, said method comprising:

(a) probing said first sample with one or more recognition means, each recognition means recognizing a different target nucleotide subsequence or a different set of target nucleotide subsequences;

(b) generating one or more output signals from said first sample probed by said recognition means, each output signal being produced from a first nucleic acid in said sample by recognition of one or more target nucleotide subsequences in said first nucleic acid by said recognition means and comprising a representation of (i) the length between occurrences of target nucleotide subsequences in said first nucleic acid, (ii) the identities of said target nucleotide subsequences in said first nucleic acid or the identities of said sets of target nucleotide subsequences among which are included the target nucleotide subsequences in said first nucleic acid, and (iii) a measure of the level/amount of the first nucleic acid in the first sample producing the output signal; and

(c) comparing each representation provided by said first sample with a second

population of second nucleic acids having different nucleotide sequences, wherein the different nucleotide sequences of said second population are known, by generating a statistical score which states the probability that one or more output signals from said first sample indicates that one or more nucleotide sequences of said first sample are either present or absent in said second sample;

whereby said indication of presence or absence provides said identification, classification, or quantitation.

25. The method of claim 24, wherein the output signals from said first sample indicate that one or more nucleotide sequences of said first sample are present in said second sample.

26. The method of claim 24, wherein the output signals from said first sample indicate that one or more nucleotide sequences of said first sample are absent in said second sample.

27. A method for identifying, classifying, or quantifying one or more nucleic acids in a first sample comprising a first population of first nucleic acids having different nucleotide sequences, said method comprising:

(a) probing said first sample with one or more recognition means, each recognition means recognizing a different target nucleotide subsequence or a different set of target nucleotide subsequences;

(b) generating one or more output signals from said first sample probed by said recognition means, each output signal being produced from a first nucleic acid in said sample by recognition of one or more target nucleotide subsequences in said first nucleic acid by said recognition means and comprising a representation of (i) the length between occurrences of target nucleotide subsequences in said first nucleic acid, (ii) the identities of said target nucleotide subsequences in said first nucleic acid or the identities of said sets of target nucleotide subsequences among which are included the target nucleotide subsequences in said first nucleic acid, and (iii) a measure of the level/amount of the first nucleic acid in the first sample producing the output signal;

(c) performing steps (a) and (b) for a second population of second nucleic acids having different nucleotide sequences, wherein the different nucleotide sequences of said second population are not known; and

(d) comparing each representation provided by said first sample with said second sample, by generating a statistical score which states the probability that one or more output signals from said first sample indicates that one or more nucleotide sequences of said first sample are either present or absent in said second sample;

whereby said indication of presence or absence provides said identification, classification or quantitation.

28. The method of claim 27, wherein the output signals from said first sample indicate that one or more nucleotide sequences of said first sample are present in said second sample.
29. The method of claim 27, wherein the output signals from said first sample indicate that one or more nucleotide sequences of said first sample are absent in said second sample.
30. The method of any of claims 18, 21, 24 or 27, wherein said statistical score is generated according to Bayes Rule of conditional probability, wherein said statistical score is related to the conditional probability that one or more second nucleic acids from said second population occurs in said first population when one or more representations of a first set of said first nucleic acids are present in said one or more output signals and no representation of a second set of first nucleic acids is present in said one or more output signals.
31. The method of any of Claims 18, 21, 24 or 27, wherein said indication that one or more nucleotide sequences of said first sample is present in said second sample is established when the conditional probability that an output signal is detected when no second nucleic acid is present in said second population is less than a predetermined limit.
32. The method of claim 31, wherein said predetermined limit is less than about 0.1.
33. The method of claim 31, wherein said predetermined limit is less than about 0.01.
34. The method of claim 31, wherein said predetermined limit is less than about 0.001.

35. The method described in any of claims 18, 21, 24 or 27, wherein the recognition means comprises one or more restriction endonucleases and the probing comprises contacting said first sample with said recognition means.
36. The method described in any of claims 18, 21, 24 or 27, wherein the recognition means comprises one or more oligonucleotides of defined sequence, and the probing comprises the steps of
- (a) contacting the first sample with the recognition means to provide an amplification mixture; and
 - (b) conducting an amplification process on said amplification mixture.
- ~~37.~~ A method for comparing one or more polypeptides in a first sample comprising a first population of polypeptides having different polypeptide sequences, with a second sample comprising a second population of polypeptides having different polypeptide sequences, said method comprising:
- (a) generating one or more output signals from said first sample and said second sample by some means; and
 - (b) comparing said output signals from said first and second samples by generating a statistical score which states that the probability that one or more output signals from the first sample indicates that one or more polypeptide sequences of said first sample are either present or absent in said second sample.
38. The method of claim 37, wherein said means is selected from the group consisting of proteolysis, mass spectroscopy, and high performance liquid chromatography.
39. The method of claim 37, wherein said statistical score is generated according to Bayes Rule of conditional probability.